

## 本周周报(9.3-9.9):

解聪

### 本周工作:

#### 1. 淘宝交易地图

##### 交易路径存在的问题:

1. 使用核密度估计可以看出某条路径的密度情况, 比如: 北京-上海的路线发货密集。因为这条路径经过: 南京, 济南, 天津很多城市, 从图中会产生如下误解: 沿线的每个城市交易量都很大。而实际情况是不是这样却无法得知。
2. 无法得知某条特定路径的信息。比如南京到济南的两城间交易信息。因为南京到济南被北京到上海的路径所覆盖。因此只能得到北京-上海这一线路的总的交易量信息, 而无法得到单条路线的信息。

##### 解决方法:

1. 改进路径分布算法。

在使用了铁路数据以后, 交易的路径被规定在铁路线路之内了(飞机线路不考虑的情况下)。

这样使得不少交易线路被重叠在了一起, 比如南京-北京的线路与济南-上海的线路。因此最直接的想法是, 除了起点终点位置不变以外, 将中间路径尽量错开布局。

##### 路径布局的目标:

- 可以考虑将铁路数据看成骨架, 而交易路径依据骨架采用某种分布, 使得用户可以看出局部某条路径信息。
- 交易路径不严格按照骨架, 交易路径又保留一定的铁路信息, 使得用户可以看出总的路径走势的信息。

##### 路径布局的好处:

- 以此来错开各条线路, 比如南京-北京与上海-济南。
- 路径可能绕开沿途的城市, 就避免了沿途所有城市都具有较大交易量的误解。比如北京-上海沿途交易量很大, 使得途径的二线城市密度值也较大, 但是它交易可能就比较小。

联想到基于骨架的边绑定, 但是这里的问题与边绑定不同之处在于:

- 这里是要将铁路路径上的重叠的边分开, 而不是将相似走向的边绑定。
- 而且这里已经没有边, 而是交易路径, 可以看成是一组边。对一组边的操作更复杂。

某种意义上与边绑定是一个相反的过程, 是否可以叫做“边分散”? 不过基于骨架的边绑定对此处很有参考价值。

##### 解决方案:

- 可以采用某种力学模型。比如: 将骨架(铁路)看成是有引力的, 边之间则具有斥力。这样的模型使得边围绕骨架分散开。问题是效率上无法可能达到实时运行。
- 直接采用概率模型计算, 问题是折线(路径)布局的算法不如力学模型直观, 同样算法可能也会很复杂。

2. 添加交互以来显示某个路径, 某些城市或某个区域的具体交易信息。比如點選两个城市高亮其之间

的交易。这样问题就虽然变简单了，但是可能效果不如上一种方法好。而且在没有交互的情况下仍然无法获得细节信息。

3.

### 本周改进:

在先前工作的基础上先尝试了将路径分开排布，本周对交易的路径的布局略作修改，如下图所示：

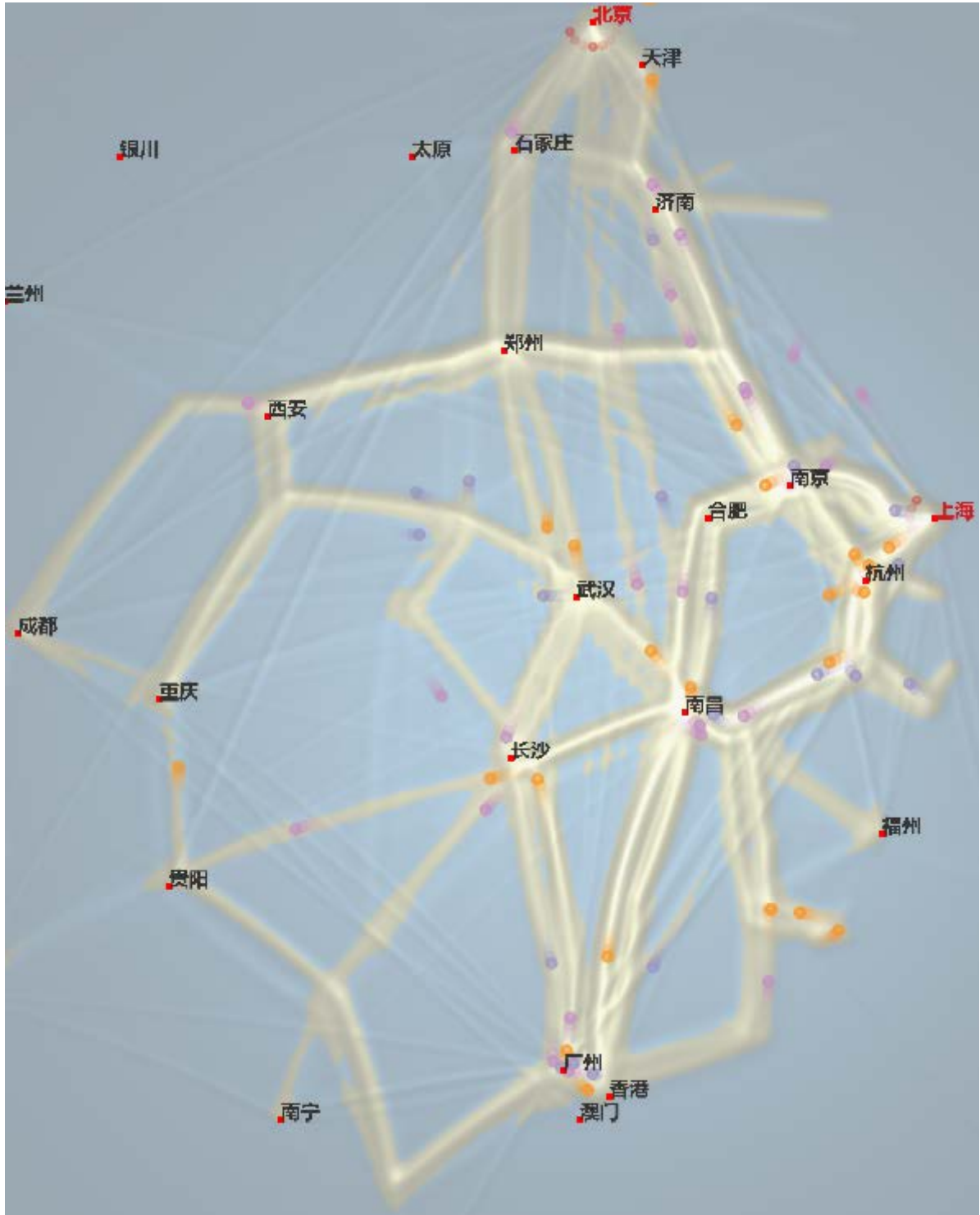


图 1 对路径布局做了小修改。图中红色的点表明各个城市的具体位置。将原来位于统一铁路线上的路径错开。比如上海到北京的交易不再严格地经过济南，因此从图中可以看出交易大多只是经过济南而已，而不是发往济南。

以上初步尝试的问题在于：由于采用不同的密度核的叠加，使得本来可以分开的特定路径信息变得模糊。因此上图的效果还不是很好，仍然看不出特定路径的信息。另外分开路径的方法也只是将路径做了相对于铁路的偏移而已，并未采用上文所述较复杂的算法，存在很大的改进之处。

附上先前结果以作对比：

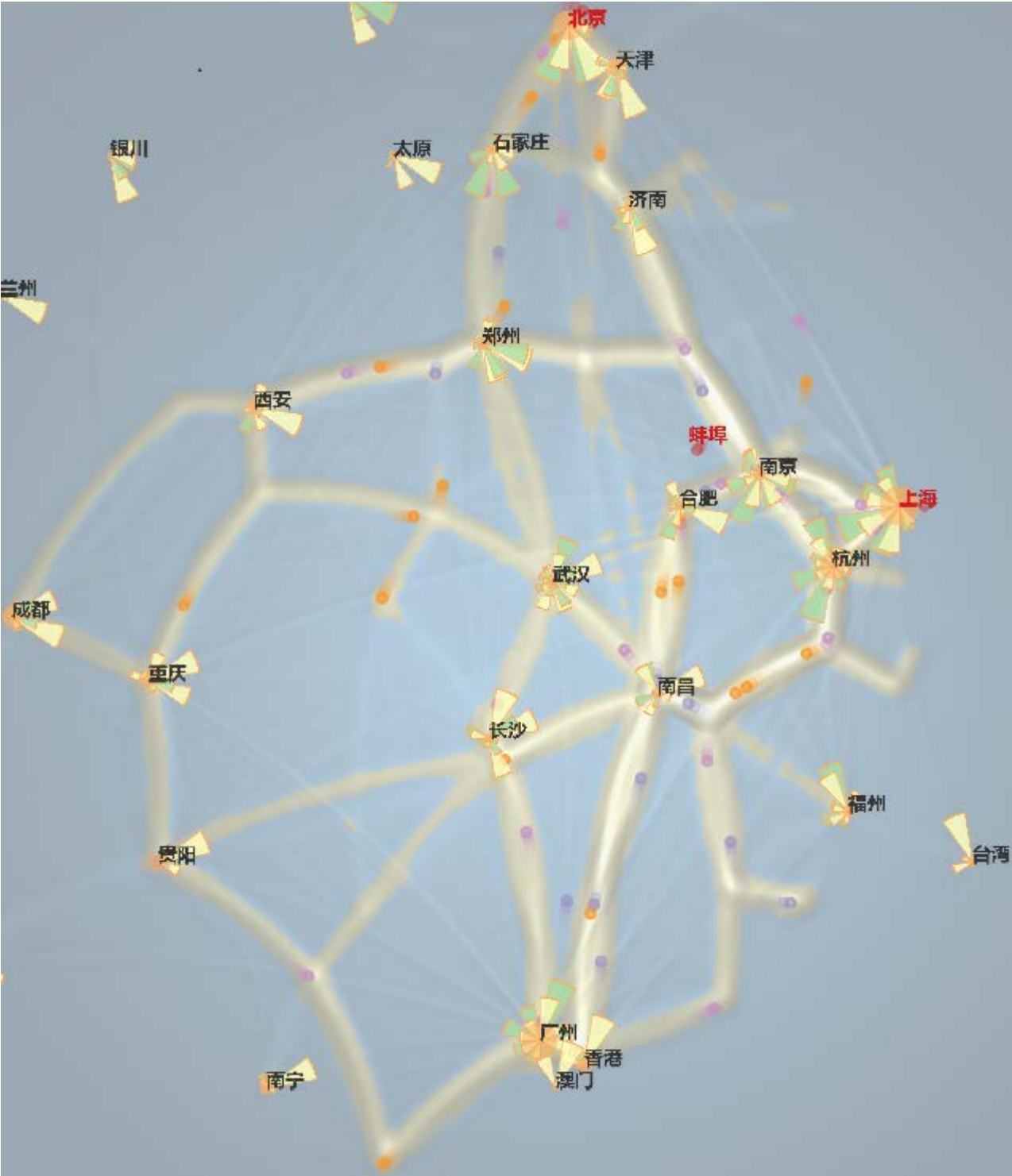


图 2 原先的结果以作对比。图中可以看出某些区域的交易比较多。比如位于京沪线上的济南，但是无从得知这些交易是路过的交易还是以济南为始、末点的交易。

本周另外对一级城市信息的可视化做了改进。按照上次讨论的一种方案做尝试。即每个城市的可视化采用相同的半径圆环，而使用颜色来区分个方向的交易量的大小。仍然分为 8 个方向，但不再使用半径大小来表示各方向之间的相对大小。

而同城的交易采用最外层的圆环及其颜色表示，颜色由浅到深表示同城交易量由小到大。

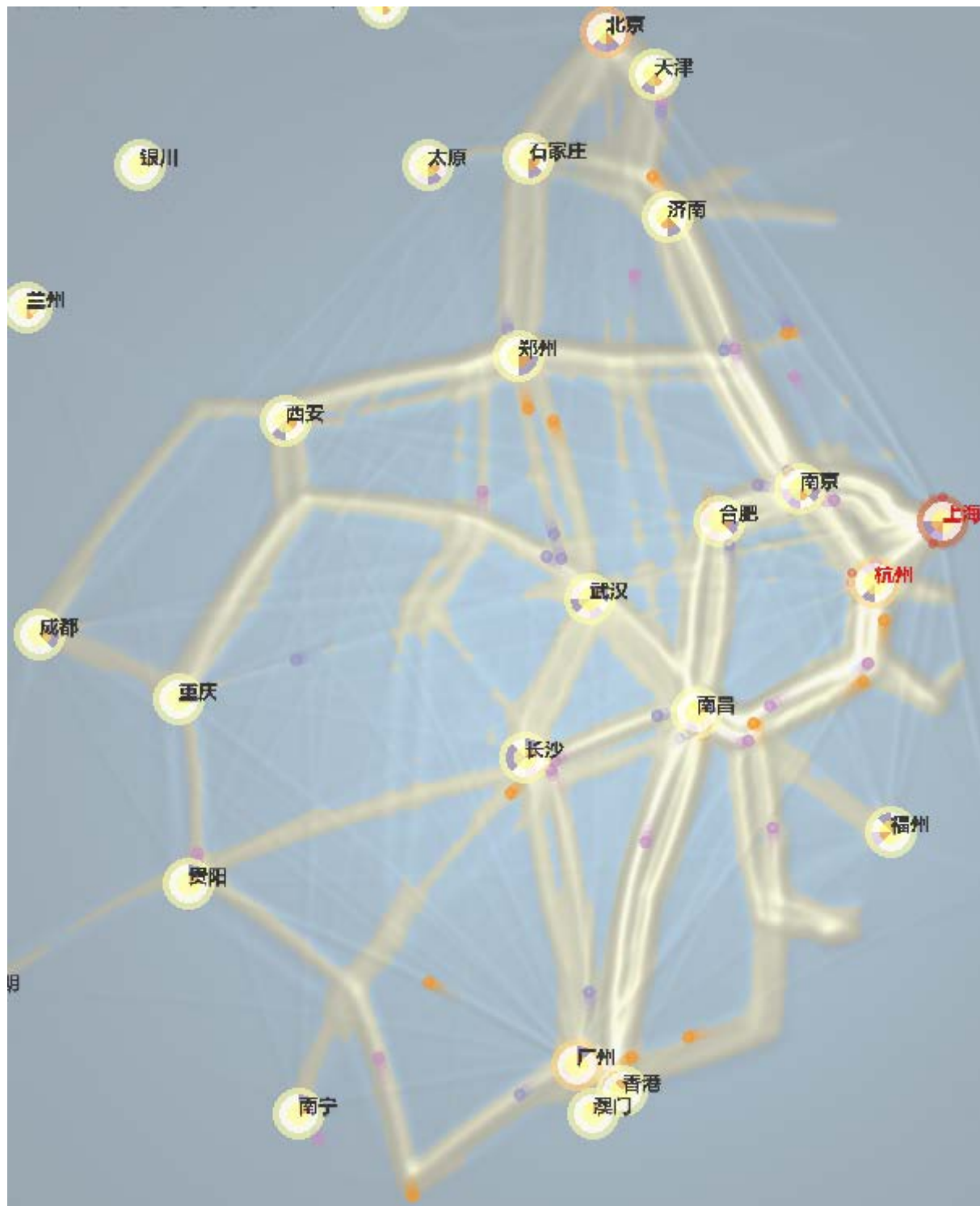


图 3 另一种一级城市可视化的尝试。最内层扇形（黄色到橙色）表示进入量大小，中间（黄色到紫色）表示发货量大小，最外层（黄色到红色）是同城交易量。三者都用颜色表明大小，颜色越深表示交易量越大。

这种方法的问题比较明显。对于交易量较小的西部城市，其颜色基本上都是浅色，表示没有交易。但是三层圆环却仍然占据了不少平面的空间。

## **2. DataV 组件**

本周三公司开会讨论了下一步，第二期的工作，包括：

- 代码结构的抽象。
- 按照发布以后的用户的反馈，添加三个基本的可视化方式：线形，柱状，饼状图。
- 对外部接口的重新修改。
- 其他一些修改。

## **3. 参加了托福考试**

下周工作：

### **1. 淘宝交易地图**

进一步按照上文的计划，探索路径的可视化方式。

### **2. DataV 组件的开发**

调整原有组件的代码结构，以及新组建的编码。

### **3. 教材**